LEVEL

SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR

BY

HIROTUGU AKAIKE

TECHNICAL REPORT NO. 40

AUGUST 1979

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

D D C

RECEIVED

OCT 29 1979

D

79 10 26 020

⑥ SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR.

By

⑩ Hirotugu/Akaike

⑨ Technical Report No. 40
August 1979
⑪

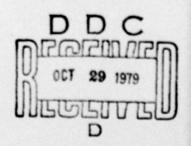⑫ 21

⑭ TR-40

⑮

PREPARED UNDER CONTRACT N00014-75-C-0442

(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

DEPARTMENT OF STATISTICS
STANFORD   UNIVERSITY
STANFORD, CALIFORNIA

D D C
RECEIVED
OCT 29 1979
D

332 580

SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR[1]

by

Hirotugu Akaike[2]

1. INTRODUCTION

Shiller [4] introduced the concept of smoothness priors to develop a
Bayesian estimator of the lag coefficients, or the sequence of impulse
responses, of a linear system. In this approach the prior preference of
the smoothness of the pattern of a set of lag coefficients is expressed
by a spherical normal distribution for the fixed order differences of
the lag coefficients. With an appropriate choice of the parameter of
the smoothness prior it is found that Shiller's distributed lag estimator
can produce results which satisfy our prior, or psychological, expecta-
tion of the smoothness of the pattern of the lag coefficients.

One practically significant problem in the application of Shiller's
estimator is the choice of the variance of the spherical normal prior
distribution. In this paper we propose a practical solution to this
problem which is obtained by maximizing the likelihood of the Bayesian
model with respect to the hyperparameter, the variance. A serious problem
about Shiller's approach is whether the assumption of smoothness of the
lag coefficients is a natural one. When the number of significant

---

[2]The Institute of Statistical Mathematics, Tokyo.

coefficients is expected to be small, which is often the case with quarterly or annual economic data, this smoothness assumption may not be quite adequate.

A natural expectation for this type of situation seems to be the smoothness of the behavior of the system characteristic in the frequency domain. The integral of the squared absolute value of the derivative of the frequency response function will then provide a measure of smoothness. This measure takes a particularly simple form in terms of the lag coefficients and leads to the definition of a non-spherical normal smoothness prior.

2. SMOOTHNESS IN FREQUENCY DOMAIN

Consider the stochastic linear system

$$y_n = \sum_{m=0}^{M} a_m x_{n-m} + w_n ,$$

(1)

where y_n, x_n, and w_n denote the output, input and error term of the system, respectively. Here $\{w_n\}$ is assumed to be a sequence of random variables which are independent of $\{x_n\}$ and are independently identically distributed as normal with mean zero and variance σ^2. The frequency response function of the output y_n to the input x_n is defined by

$$A(f) = \sum_{m=0}^{M} a_m \exp(-i2\pi mf) .$$

(2)

We propose to measure the smoothness of the frequency response function by

$$R = \int \left| \frac{dA(f)}{df} \right|^2 df \; ,$$

where $\left| dA(f)/df \right|$ denotes the absolute value of $dA(f)/df$ and the integral is over $-0.5 \leq f \leq 0.5$. As is obvious from the definition, a large value of R means a rapid change of the system characteristic against the change of the frequency characteristic of the input. Thus, actually, R is a measure of unsmoothness of $A(f)$. Since it holds that

$$\frac{dA(f)}{df} = \sum_{m=0}^{M} (-12\pi m) a_m \exp(-12\pi mf) \; ,$$

we have

(3) $$R = (2\pi)^2 \sum_{m=1}^{M} m^2 a_m^2 \; .$$

3. A SMOOTHNESS PRIOR

As is discussed in Akaike [2] we consider the minimization of the quantity

$$\sum_{n=1}^{N} \left(y_n - \sum_{m=0}^{M} \alpha_m x_{n-m} \right)^2 + \lambda^2 \sum_{m=1}^{M} m^2 \alpha_m^2 \; .$$

By a proper choice of the constant λ, this minimization leads to the determination of α_m with small sum of squares of the residuals and also with smooth behavior of the frequency response function.

The minimization problem is transformed into a statistical problem by embedding it into the maximization problem of

$$(4) \qquad \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}\left(y_n - \sum_{m=0}^{M}\alpha_m x_{n-m}\right)^2\right]\exp\left(-\frac{\lambda^2}{2\sigma^2}\sum_{m=1}^{M}m^2\alpha_m^2\right).$$

For a given σ, the first factor of (4) obviously represents the likelihood of the α_m's. The second factor represents our prior preference on α_m's in the form of a normal distribution. Our preference on α_0 and σ is not specified in (4) and we assume the improper prior

$$\frac{d\alpha_0}{\sigma}\frac{d\sigma}{\sigma}$$

for α_0 and σ. This choice is based on the consideration that it is an ignorance prior of the parameters of a normal distribution $N(c\alpha_0, \sigma)$, where c is a nonzero constant. For the justification of the use of this ignorance prior distribution for its impartial performance, see Akaike [1].

Our Bayesian model is thus defined by the data distribution

$$(5) \qquad p(y|\alpha,\sigma) = \left(\frac{1}{2\pi\,\sigma^2}\right)^{\frac{N}{2}}\exp\left(-\frac{1}{2\sigma^2}\|y - X\alpha\|^2\right)$$

and the (improper) prior distribution

$$(6) \qquad p(\alpha, \sigma | \lambda) = \left(\frac{1}{2\pi \sigma^2}\right)^{\frac{M+1}{2}} |\lambda^2 DD'|^{\frac{1}{2}} \exp\left(-\frac{\lambda^2}{2\sigma^2} \alpha' D' D\alpha\right)\left(\frac{1}{\sigma}\right) ,$$

where $y = (y_1, y_2, \ldots, y_N)'$, $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_M)'$ and X and D are respectively $N \times (M+1)$ and $M \times (M+1)$ matrices defined by

$$X = \begin{bmatrix} x_1 & x_0 & \cdots & x_{-M+1} \\ x_2 & x_1 & \cdots & x_{-M+2} \\ \vdots & \vdots & & \vdots \\ x_N & x_{N-1} & \cdots & x_{N-M} \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & M \end{bmatrix} ,$$

and where $\|y\|$ denotes the Euclidean norm of a vector y and $|A|$ denotes the determinant of a matrix A and ' denotes transposition.

4. DETERMINATION OF THE HYPERPARAMETER

In his original work Shiller [4] demonstrated by a numerical example the insensitivity of his estimate to the choice of the hyperparameter of his smoothness prior. This hyperparameter corresponds to λ of (6). We often consider this insensitivity as a proof of the robustness of the Bayesian model and take it as a justification for an arbitrary choice of the value of the hyperparameter. Nevertheless, it is only when we know the true lag coefficients that we can see that the estimates thus obtained do approximate the true coefficients. Thus it is desirable to have an objectively defined procedure for the selection of the value of the hyperparameter.

If the prior distribution $p(\alpha,\sigma|\lambda)$ is proper, the likelihood of the hyperparameter λ with respect to the data y is defined by

$$(7) \qquad p(y|\lambda) = \int p(y|\alpha,\sigma)\, p(\alpha,\sigma|\lambda)\, \alpha\,d\alpha\,d\sigma .$$

A reasonable choice of λ is realized by maximizing the likelihood $p(y|\lambda)$. This procedure for the determination of a hyperparameter is called the method of Type II maximum likelihood by Good [3]. In our case the distribution $p(\alpha,\sigma|\lambda)$ given by (6) contains an improper component. Nevertheless, for each particular situation, the improper component can be considered as an approximation to a very widely dispersed proper prior distribution. Thus we take (7) as the definition of the likelihood of the prior distribution defined by (6) and propose the use of λ which maximizes $p(y|\lambda)$. Obviously it would be more desirable to develop an ignorance-type prior of λ , but at present we leave this as a subject of further research.

For the present Bayesian model defined by (5) and (6) we have

$$p(y|\alpha,\sigma)\, p(\alpha,\sigma|\lambda) = \left(\frac{1}{2\pi\,\sigma^2}\right)^{\frac{N+M+1}{2}} |\lambda^2\, DD'|^{\frac{1}{2}} \left(\frac{1}{\sigma}\right)$$

$$\cdot \exp\left[-\frac{1}{2\sigma^2}\,(\alpha-\alpha_0)'(X'X + \lambda^2\, D'D)(\alpha-\alpha_0)\right] \exp\left[-\frac{1}{2\sigma^2}\, S(\lambda)\right] ,$$

where $\alpha_0 = (X'X + \lambda^2\, D'D)^{-1}\, X'y$ and $S(\lambda) = \|y\|^2 - \alpha_0'(X'X + \lambda^2 D'D)\alpha_0$. Thus we have

6

$$p(y|\lambda) = (\tfrac{1}{2})(\tfrac{1}{\pi})^{\frac{N}{2}} \Gamma(\tfrac{N}{2}) \; S(\lambda)^{-\frac{N}{2}} \; |X'X + \lambda^2 D'D|^{-\frac{1}{2}} \; |\lambda^2 DD'|^{\frac{1}{2}} \; .$$

By applying the formal Bayes procedure we get the posterior mean of α and σ^2. They are given by α_0 and $S(\lambda)/(N-2)$, respectively. Computationaly, α_0 and $S(\lambda)$ can be obtained directly by using the relation

(8)
$$S(\lambda) = \operatorname*{Min}_{\alpha} \|y\star - X\star\alpha\|^2 = \|y\star - X\star\alpha_0\|^2 \; ,$$

where

$$y\star = \begin{array}{c} \uparrow \\ N \\ \downarrow \\ \uparrow \\ M \\ \downarrow \end{array} \begin{bmatrix} y \\ \\ 0 \end{bmatrix} \quad \text{and} \quad X\star = \begin{array}{c} \uparrow \\ N \\ \downarrow \\ \uparrow \\ M \\ \downarrow \end{array} \begin{bmatrix} X \\ \\ \lambda D \end{bmatrix} \; .$$

The search for the λ which maximizes $p(y|\lambda)$ may practically be limited to a set of finite discrete values of λ. Since we are familiar with the log likelihood ratio test statistic, we replace $p(y|\lambda)$ by $(-2) \log p(y|\lambda)$ and try to minimize it. Ignoring an additive constant, $(-2) \log p(y|\lambda)$ is given by

(9)
$$L(\lambda) = N \log S(\lambda) + \log|X'X + \lambda^2 D'D| - \log|\lambda^2 DD'| \; .$$

Akaike [2] reached the same procedure for the determination of λ by

using the maximum likelihood estimate of σ^2 instead of developing the improper prior distribution of σ .

5. NUMERICAL EXAMPLES

To check the performance of our procedure it was applied to the data generated by the relation

$$y_n = 1.2 \, x_n - 0.6 \, x_{n-1} + 0.4 \, x_{n-2} + w_n \, ,$$

where the input x_n was the same as that used in the second example of Shiller [4] and w_n was normal with mean zero and $\sigma = 0.05$. The length N of the time series was 40. The Bayesian model was developed with the highest lag $M = 19$. The search for the minimum of $L(\lambda)$ was limited to the values $\lambda = 10 \times 2^k$ ($k = -9,-8,\ldots,-1$) and the minimum was attained with $k = -5$. The corresponding estimate α_0 is given in Table 1 along with the true parameter and the estimate obtained by assuming the Shiller's smoothness prior for the second order differences $(\alpha_{i+2} - \alpha_{i+1}) - (\alpha_{i+1} - \alpha_i)$ ($i = 0,1,\ldots,M-2$) . The estimate denoted by Shiller was obtained by putting

$$D = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & 1 & -2 & 1 \end{bmatrix}$$

in (6) and then minimizing the corresponding $L(\lambda)$ for $\lambda = 10 \times 2^k$ ($k = -5,-4,\ldots,10$). The minimum was attained at $k = 10$. Further increase of k would produce estimates smoother than that given in Table 1.

8

Table 1

Comparison of estimates. α_0 denotes the estimate obtained by the smoothness prior in the frequency domain. Shiller denotes the estimate obtained by assuming Shiller's prior for the second order differences.

m	0	1	2	3	4
True	1.2	-0.6	0.4	0	0
α_0	1.137	-0.440	0.230	0.050	0.017
Shiller	0.203	0.186	0.170	0.154	0.137

m	5	6	7	8	9
True	0	0	0	0	0
α_0	0.012	-0.000	0.003	0.001	-0.002
Shiller	0.121	0.105	0.089	0.072	0.056

m	10	11	12	13	14
True	0	0	0	0	0
α_0	0.001	0.003	-0.000	-0.000	-0.003
Shiller	0.040	0.023	0.007	-0.009	-0.025

m	15	16	17	18	19
True	0	0	0	0	0
α_0	-0.005	-0.006	-0.001	0.004	0.001
Shiller	-0.042	-0.058	-0.074	-0.090	-0.048

The result given in Table 1 demonstrates the necessity of great care in applying Bayesian models. Although the Shiller estimate is definitely in accordance with the prior preference of smoothness it is giving a definitely biased image of the true impulse response sequence. The sum of squared errors of an estimate α is defined by $SSE = \|\alpha - a\|^2$, where a denotes the true parameter. The values of SSE are 0.06 and 1.77 for our estimate and the Shiller estimate, respectively. Another measure of inaccuracy of an estimate α is defined by the sum of squared errors of regression defined by

$$SSER = \|X(\alpha - a)\|^2 .$$

The values of SSER are 0.0008 and 0.1955 for our estimate and the Shiller estimate, respectively. Thus the bias introduced by assuming Shiller's smoothness prior seems to be producing significant influence on the estimation of regression. The result shown in Table 1 is the first one obtained in a series of three experiments all of which produced similar results.

One might argue that the situation will be reversed when the present prior is applied to the case where Shiller's prior is appropriate. To check this possibility the estimates based on the two priors were obtained for a sample from the model treated in Shiller's second example. The result is shown in Table 2. For the purpose of comparison the ordinary least squares estimate, denoted by LS, is also included. The result demonstrates the extremely good performance of the Shiller estimate. This result is already reported in Akaike [2]

10

Table 2

Comparison of estimates. α_0 and Shiller are as defined in Table 1. LS denotes least squares estimate.

m	0	1	2	3	4
True	.000	.000	.001	.004	.018
α_0	.004	-.005	-.011	.031	.037
Shiller	-.009	-.003	.004	.009	.017
LS	-.010	-.021	-.045	.037	.078

m	5	6	7	8	9
True	.054	.130	.242	.352	.399
α_0	-.006	.176	.208	.372	.378
Shiller	.051	.134	.242	.345	.395
LS	-.074	.255	.113	.462	.334

m	10	11	12	13	14
True	.352	.242	.130	.054	.018
α_0	.366	.260	.107	.055	.033
Shiller	.362	.257	.134	.052	.012
LS	.359	.329	.046	.072	.042

m	15	16	17	18	19
True	.004	.001	.000	.000	.000
α_0	-.016	-.020	.021	.012	-.003
Shiller	-.011	-.015	.006	.035	-.018
LS	-.018	-.050	.065	-.008	-.008

11

and is considered as a demonstration of the performance of our procedure for the selection of λ . The present result also shows that our estimate does not show any significant bias and that it is closer to the Shiller estimate than to the least squares estimate. The sum of squared errors of the estimate, SSE, is 0.0117 for our estimate, while it is 0.0026 for the Shiller estimate. Nevertheless, the sum of squared errors of regression, SSER, is 0.00047 for our estimate, while it takes a larger value 0.00051 for the Shiller estimate. This suggests that the use of the present prior in the situation where Shiller's prior is more appropriate may not necessarily induce significant loss in terms of the accuracy of regression.

6. DISCUSSION

The purpose of the present paper has been two-fold. The first is to introduce a new smoothness prior which might find a wider applicability than the Shiller's original smoothness prior. The second is to propose a practically useful objectively defined procedure for the choice of the hyperparameter. The numerical results given in the preceding section show that these objectives are attained fairly well.

One of the most significant results is the numerical demonstration of the possible significant bias due to the Bayesian modeling. In the Bayesian approach we specify our psychological expectation in the form of a prior distribution and proceed to get a result which satisfies this expectation. Our numerical results dramatically demonstrated that this satisfaction of psychological expectation may be quite

deceptive and misleading. Thus we see that no unique Bayesian model can be assured of a success in its practical application. The only reasonable procedure would then be to propose several possible Bayesian models and compare the likelihoods of these models, and, if necessary and feasible, develop a larger Bayesian model using these models.

Certainly the smoothness prior in the frequency domain introduced in this paper is not quite satisfactory to cover every practical situation. When some definite delay in the response is conceivable, i.e., when $\alpha_0, \alpha_1, \ldots, \alpha_d$ are expected to be close to zero for some $d(<M)$, it may be more appropriate to assume a prior distribution over the possible delays d and, for each d, replace the definition of x_n by x_{n-d-1}. Practically it may be better to replace $A(f)$ in the definition (3) of R by $\exp(i2\pi f\delta) A(f)$ defined with some properly chosen integer δ. This replaces R of (3) by

$$R_\delta = (2\pi)^2 \sum_{m=0}^{M} (m-\delta)^2 a_m^2 ,$$

which clearly demonstrates the effect of the choice of δ. Since we do not know which value of δ we should use, we will assume a prior distribution over a possible set of values of δ, such as $(-1, 0, 1, 2, 3)$, and define the "likelihood" of each model by

$$\ell_\delta = \exp[-\frac{1}{2} \max L_\delta(\lambda)] ,$$

where $\max L_\delta(\lambda)$ denotes the maximum of $L(\lambda)$ defined by (9) for a

13

specific choice of δ. With a proper modification of D in (6) we may even assume δ to be half integers, such as $-0.5, 0.5, 1.5, \ldots$. The practical utility of this type of approach needs further investigation.

What is practically more important than the refinement of the Bayesian model is the recognition of the limited applicability of the basic model (1) to economic time series. This is due to the fact that there is usually a feedback from the output y_n to the input x_n. Only in the special situation where this feedback is negligible can we expect the use of the present model. Thus the result reported in this paper must be considered as only the second step following the first step of the original contribution of Shiller in making the Bayesian modeling a practical procedure. Further elaboration of the basic model is definitely necessary to make the procedure widely applicable to the analysis of complex economic time series.

REFERENCES

[1] Akaike, Hirotugu: "A New Look at the Bayes Procedure", Biometrika, 65 (1978), 53-59.

[2] Akaike, Hirotugu: "Likelihood and the Bayes Procedure", A paper presented at the International Meeting on Bayesian Statistics, May 28 - June 2, 1979, Valencia, Spain. To be published in Trabajos Estadistica.

[3] Good, Irving J.: The Estimation of Probabilities. Cambridge: M.I.T. Press, 1965.

[4] Shiller, Robert J.: "A Distributed Lag Estimator Derived From Smoothness Priors", Econometrica, 41 (1973), 775-788.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.

2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.

3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.

4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.

5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.

6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.

7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.

8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.

9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.

10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.

11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.

12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.

13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.

14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.

15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.

16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

17. "General Exponential Models for Discrete Observations,"
Steffen L. Lauritzen, May, 1974.

18. "On the Interrelationships among Sufficiency, Total Sufficiency and
Some Related Concepts," Steffen L. Lauritzen, June, 1974.

19. "Statistical Inference for Multiply Truncated Power Series Distributions,"
T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models
in the Time and Frequency Domains," T. W. Anderson, June 1975.

21. "Asymptotic Properties of Some Estimators in Moving Average Models,"
Raul Pedro Mentz, September 8, 1975.

22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
Mituaki Huzii, February 1976.

23. "Estimating Means when Some Observations are Classified by Linear
Discriminant Function," Chien-Pai Han, April 1976.

24. "Panels and Time Series Analysis: Markov Chains and Autoregressive
Processes," T. W. Anderson, July 1976.

25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson,
September 1976.

26. "The Recurrence Classification of Risk and Storage Processes,"
J. Michael Harrison and Sidney I. Resnick, September 1976.

27. "The Generalized Variance of a Stationary Autoregressive Process,"
T. W. Anderson and Raul P.Mentz, October 1976.

28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
Javad Behboodian, October 1976.

29. "Identification of Parameters by the Distribution of a Maximum
Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.

30. "Discrimination Between Stationary Guassian Processes, Large Sample
Results," Will Gersch, January 1977.

31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
Christine M. Waternaux, October 1977.

32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a
Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

33. "Canonical Correlations with Respect to a Complex Structure,"
     Steen A. Andersson, July 1978.

34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and
     I. Olkin, July 1978.

35. " Maximum likelihood Estimation for Vector Autoregressive Moving
     Average Models," T. W. Anderson, July 1978.

36. "Maximum likelihood Estimation of the Covariances of the Vector Moving
     Average Models in the Time and Frequency Domains," F. Ahrabi, August 1978.

37. "Efficient Estimation of a Model with an Autoregressive Signal with
     White Noise," Y. Hosoya, March 1979.

38. "Maximum Likelihood Estimation of the Parameters of a Multivariate
     Normal Distribution,"T.W. Anderson and I. Olkin, July 1979.

39. "Maximum Likelihood Estimation of the Autoregressive Coefficients and
     Moving Average Covariances of Vector Autoregressive Moving Average Models,"
     Fereydoon Ahrabi, August 1979.

40. "Smoothness Priors and the Distributed Lag Estimator," Hirotugu Akaike,
     August 1979.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 40	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SMOOTHNESS PRIORS AND THE DISTRIBUTED LAG ESTIMATOR		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) HIROTUGU AKAIKE		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPARTMENT OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIFORNIA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH STAT. & PROBABILITY PROGRAM CODE 436 ARLINGTON, VA. 22217		12. REPORT DATE AUGUST 1979
		13. NUMBER OF PAGES 14
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Smoothness priors, distributed lag, frequency response function, Bayesian estimation, hyperparameter, Type II maximum likelihood, Schiller's prior.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE OTHER SIDE

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-014-6601

20. Technical Report No. 40

ABSTRACT

Shiller's distributed lag estimator based on a smoothness prior demonstrates the potential of the Bayesian approach to statistical model building. Nevertheless, when the number of significant lag coefficients is small the assumption of smoothness of the pattern of the lag coefficients may not be appropriate. In this paper, to cover such a situation, the smoothness is assumed for the behavior of the coefficients viewed in the frequency domain. This definition leads to a smoothness prior with a particularly simple form. Numerical result shows that the estimator based on this smoothness prior produces good estimates of the lag coefficients where Shiller's prior produces highly biased estimates. It is also observed that the new estimator produces reasonable results even when the Shiller's prior is more appropriate. The danger of introducing a bias by assuming a Bayesian model is stressed in the discussion.